

1

2

3 **Skill of the MJO and Northern Hemisphere Blocking**
4 **in GEFS Medium-Range Reforecasts**

5

6

7

8 Thomas M. Hamill and George N. Kiladis
9

10

11

12 *NOAA Earth System Research Lab*
13 *Physical Sciences Division*
14 *Boulder, Colorado, USA*

15

16

17

18

19

20

21

22

Submitted to *Monthly Weather Review*

23

24

25

26

27

28

29

30

31

Corresponding author address:

32

33

Dr. Thomas M. Hamill

34

NOAA Earth System Research Lab, Physical Sciences Division
R/PSD1, 325 Broadway

35

Boulder, Colorado USA 80305-3328

36

e-mail: tom.hamill@noaa.gov

37

phone: +1 (303) 497-3060 telefax +1 (303) 497-6449

38

39

40 ABSTRACT
41
42 Forecast characteristics of Northern Hemisphere atmospheric blocking and
43 the MJO were diagnosed using an extensive time series (Dec-Jan-Feb 1985-2012) of
44 daily medium-range ensemble reforecasts based on a version of the NCEP Global
45 Ensemble Forecast System (GEFS).
46 For blocking : (a) inter-annual variability of analyzed blocking frequency was
47 quite large; (b) the GEFS slightly under-forecasted blocking frequency at longer
48 leads in the Euro-Atlantic sector; (c) predictive skill of actual blocking was
49 substantially smaller than its perfect-model skill; (d) block onset and cessation were
50 forecast less well than overall blocking frequency; (e) there was substantial
51 variability of blocking skill between half-decadal periods; and (f) the reliability of
52 probabilistic blocking forecasts degraded with increasing lead time.
53 For the MJO: (a) forecasts of strong Indian Ocean MJOs propagated too slowly,
54 especially the component associated with outgoing longwave radiation (OLR), i.e.,
55 convection; (b) tropical precipitation was greatly over-forecast at early lead times;
56 (c) the ensemble predictions were biased and/or under-dispersive, manifested in U-
57 shaped rank histograms of MJO indices. Magnitude forecasts were especially U-
58 shaped. (d) MJO correlation skill was larger for its wind than for its OLR component,
59 and was larger for the higher-amplitude MJO events; (e) there was some half-
60 decadal variability in skill; (f) probabilistic skill of the MJO forecast was modest, and
61 skill was larger when measured relative to climatology than when measured
62 relative to a lagged persistence forecast.

63 For longer-lead forecasts, the GEFS demonstrated little ability to replicate the

64 changes in blocking frequency due to a strong MJO that were noted in analyzed data.

65

66 **1. Introduction.**

67 Medium-range predictability and forecastability¹ of low-frequency modes of
68 atmospheric variability can be evaluated more readily with a long time series of
69 forecasts. In this manuscript we evaluate two such modes of variability, Northern
70 Hemispheric atmospheric blocking and the Madden-Julian Oscillation (MJO; Madden
71 and Julian 1971) and their interactions using an extensive set of global medium-
72 range ensemble reforecasts.

73 Both the MJO and blocking occur somewhat infrequently at a particular
74 location. Using the blocking definition of Tibaldi and Molteni (1990), averaged over
75 several decades the Northern Hemisphere wintertime blocking frequency ranges
76 from approximately 2% to 22%, depending on the longitude. Blocks are most
77 common over the eastern Atlantic Ocean and western Europe, with a secondary
78 frequency maximum in the central Pacific Ocean. Blocks can be persistent, leading
79 to both long periods of rather similar, and often high-impact weather. For the slowly
80 moving, large-scale, equatorially trapped tropical convective envelopes and
81 associated wind perturbations known as the MJO, a given year may produce only a
82 couple coherent, high-amplitude MJO events, although this number will vary
83 depending on the definition of the MJO (Straub 2013).

¹ *Predictability* here refers to an intrinsic property of the atmospheric process. It measures the time scale at which two initially similar but not identical perturbed initial conditions subject to the actual atmosphere's dynamics will become as different as random draws of atmospheric states. While numerical models are often used to estimate the unknown predictability, the predictability is not a forecast property, but a property of the atmosphere. *Forecastability*, in contrast, indicates the ability of the forecast model to provide guidance that the user will judge to have some value. Forecastability may be evaluated with many different metrics.

84 Evaluating the predictability and forecastability of blocking and the MJO are
85 challenging given the infrequency and the temporal continuity of these events, the
86 latter of which reduces the effective sample size (Wilks 2006, eq. 5.12). If one is
87 interested, for example, in the ability of the model to forecast block onset for a
88 particular longitude band, a three-month period may have, say, 15 days with
89 blocked conditions, but most likely those 15 days occurred in 1-3 persistent
90 blocking events. Similar issues occur with the MJO. The interaction of two such
91 events is even more difficult to evaluate with limited samples, e.g., evaluating the
92 change in blocking frequency related to a large forecast MJO event in the Indian
93 Ocean. A large sample size could be provided from a reforecast, i.e., a multi-year or
94 preferably a multi-decadal sample of forecasts from a fixed forecast model and
95 assimilation system.

96 Extensive reforecasts (hindcasts) often haven't been available to facilitate
97 such studies. For blocking, the most comprehensive recent study was by Jung et al.
98 (2012), which used hindcasts to document blocking frequency in extended-range
99 simulations from the European Centre for Medium-Range Weather Forecasts
100 (ECMWF) model at various resolutions. The authors found that Euro-Atlantic
101 sector blocking frequency was generally more under-forecast with lower-resolution
102 models. Older studies included the Watson and Colucci (2002) study of Northern
103 Hemispheric wintertime blocking using data from the operational NCEP global
104 spectral model from 1995-1998, and Mauritzen and Källén (2004), who studied
105 blocking in the ECMWF system during the Northern Hemispheric 2000-2001 winter.
106 Both studies found too few blocks in the forecast. Pelly and Hoskins (2003a), using

107 a potential vorticity-based method of defining blocks, evaluated the ECMWF output
108 for a year of data beginning on 1 August 2001, a period spanning several model
109 changes. Though they also found blocking frequency was under-forecast, there was
110 positive skill in the probabilistic forecasts of blocks out to 10 days, and they found
111 block onset was better forecast than block cessation.

112 There have been more recent studies of the MJO than for blocking, for
113 seasonal simulations at least. The MJO is being actively studied in part because it
114 affects monsoon (Yasunari 1979) and tropical cyclone variability (Maloney and
115 Hartmann 2000ab). The MJO also can excite extra-tropical Rossby wave trains
116 (Knutson and Weickmann 1987, Jones et al. 2004, Weickmann and Berry 2009) and
117 can interact with mid-latitude, low-frequency modes of variability such as the North
118 Atlantic Oscillation (L'Heureux and Higgins 2008, Cassou 2008, Lin et al. 2009). The
119 MJO is often poorly forecast, and there is some evidence that an improved MJO
120 forecast may result in improved mid-latitude forecasts (Ferranti et al. 1990, Vitart
121 and Molteni 2010).

122 The interaction of the MJO with the mid-latitude flow and its forecastability
123 has been an active area of investigation (Liebmann and Hartmann 1984, Kiladis and
124 Weickmann 1992, Hendon et al. 2000; Riddle et al. 2012). Since the MJO represents
125 the variability at time scales of 30-70 d⁻¹, (e.g., Waliser et al. 2009), it has been
126 common to examine numerical simulations to lead times of several months and to
127 leverage hindcasts to provide large-enough samples. Examples of studies with
128 seasonal forecast models and hindcasts include Hendon et al. (2000), Lin et al.
129 (2008), Seo et al. (2009), Kim et al. (2009), Vitart and Molteni (2010), Gottschalck et

130 al. (2010), Kang and Kim (2010), Jia et al. (2010), and Crueger et al. (2013). For the
131 medium range (here, roughly +3 to +16 days lead time), the literature on MJO
132 forecast evaluation is sparser. Some verification statistics were calculated from
133 THORPEX Interactive Grand Global Ensemble (TIGGE; Bougeault et al. 2010) data by
134 Matsueda and Endo (2011).

135 Recently, NOAA scientists created an extensive global ensemble reforecast
136 data set (Hamill et al. 2013) using the version of the NCEP Global Ensemble Forecast
137 System (GEFS) that was operational in 2012-2013. This data set was created to
138 facilitate the diagnosis and statistical correction of systematic forecast errors in
139 medium-range ensemble forecasts, thereby improving GEFS guidance. In this article,
140 we demonstrate an important ancillary purpose, showing how the extensive
141 reforecasts facilitate the diagnosis of errors in low-frequency modes of variability,
142 modes that can have a profound impact on intra-seasonal forecast skill. Specifically,
143 we will examine the usefulness of the reforecasts for examining the predictability
144 and forecastability of blocking and the MJO as well as their inter-relationships. The
145 over-arching hypothesis, which as we will show is easily disproved, is that the
146 ensemble prediction system well represents the evolution of forecast uncertainty of
147 these phenomena, i.e., the ensemble forecasts and observed can be considered
148 random draws from the same underlying distribution.

149 Below, section 2 briefly describes the data set and the methods for forecast
150 evaluation. Section 3 provides results, while section 4 provides a discussion and
151 conclusions.

152

153 **2. Data and Methods**

154 *a. Description of the data sets.*

155 Unless noted otherwise, reforecast data for Dec-Jan-Feb 1985-2012 period
156 was used in this study. The global ensemble reforecast data was more completely
157 described in Hamill et al. (2013). Briefly, this reforecast data set is based on the
158 2012 version of the NCEP Global Ensemble Forecast System (GEFS). An 11-member
159 retrospective ensemble forecast was generated to +16 days lead for every day at 00
160 UTC from 1 December 1984 to the current date. Consistent with the operational
161 GEFS, the model resolution was T254L42 to day +8 (~40-km grid spacing at 40°
162 latitude and 42 levels). Starting at day +7.5 and extending to day +16, the
163 reforecasts were conducted at the reduced resolution of T190L42 (~54-km grid
164 spacing). Through 20 February 2011, the Climate Forecast System Reanalysis
165 (CFSR; Saha et al. 2010) provided the control initialization and verification.
166 Thereafter, the operational grid-point statistical interpolation (GSI; Kleist et al.
167 2009) procedure was used, which was updated to a hybrid variational-ensemble
168 data assimilation approach (Hamill et al. 2011) on 22 May 2012. Additional
169 ensemble member perturbed initial conditions were generated using the ensemble
170 transform with rescaling approach of Wei et al. (2008). See Hamill et al. (2013) for
171 more details on the data set, including a description of the extensive amount of data
172 that is available for fast-access download. In this study, data interpolated to a 1-
173 degree grid was used.

174 For examination of associated tropical precipitation forecasts, 1-degree
175 Global Precipitation Climatology Project (GPCP, Huffman et al. 2001) data was used.
176 This GPCP data was available only from 1997 – current.

177

178 *b. Blocking and MJO definitions.*

179 For blocking, though there have been some modern alternatives (e.g., Pelly
180 and Hoskins 2003b, Barnes et al. 2012), here blocking was defined directly
181 following the simple method of Tibaldi and Molteni (1990) based on 500 hPa
182 geopotential heights. Conditional climatologies of blocked and unblocked 500 hPa
183 height patterns at the international dateline are shown in Fig. 1. Given the
184 increased frequency of blocking in the Euro-Atlantic and Pacific regions (Fig. 2a),
185 our analysis of blocking forecastability and predictability was limited to two sectors,
186 the “Euro-Atlantic” sector, from 45° W longitude to 45° E longitude, and the “Pacific”
187 sector, from 140° E longitude to 130° W longitude.

188 For MJO analysis, a now-standard (Gottschalk et al. 2010) method was used
189 to compute the projections onto the two leading empirical orthogonal functions
190 (EOFs) of MJO variability (Wheeler and Hendon 2004; hereafter WH04). The
191 respective EOFs are commonly known as “RMM1” and “RMM2.” The EOF structures
192 associated with RMM1 and RMM2 were taken directly from the real-time MJO web
193 site, <http://cawcr.gov.au/staff/mwheeler/maproom/RMM/>. The EOFs were
194 computed from a time series of longitudinal arrays of filtered anomalies in the 200
195 and 850 hPa zonal winds and outgoing longwave radiation (OLR), which were
196 averaged in the band from 15° S to 15° N latitude and normalized by their variances

197 over this latitude band and all longitudes. This filtering also mostly removed inter-
198 annual variability and the projection onto the El Niño/Southern Oscillation (ENSO).
199 The processing of the reforecast data to calculate the projections of forecast data
200 onto RMM1 and RMM2 generally followed the procedure outlined by WH04, with
201 the following exception. To attempt to remove the effects of ENSO, the projection of
202 the filtered data onto the Niño 3.4 index anomalies was removed (Trenberth 1997),
203 as opposed to the “SST1” index of Drosdowsky and Chambers (2001) cited in WH04.
204 The filtering did remove the leading three harmonics of the annual cycle and the
205 mean of the previous 120 days, as in WH04. Our procedure, when applied to the
206 control initial condition from the CFSR, produced a time series of projections that
207 correlated at ~ 0.87 to the time series produced by WH04 with NCEP-NCAR
208 reanalyses (Kalnay et al. 1996) data (not shown). The discrepancies are most likely
209 due to the reduced quality of the NCEP-NCAR reanalysis relative to the CFSR
210 analysis, and the use of observed OLR in the RMMs calculated from NCEP-NCAR data
211 and reanalysis OLR from the CFSR.

212

213 *c. Methods of forecast evaluation.*

214 Many methods were used to evaluate blocking and MJO forecast skill. For
215 blocking, a Brier Skill Score was used that measured the skill of the ensemble
216 forecast’s ability to set forecast probability of blocked conditions. The Brier Score of
217 the forecast and climatology were computed separately for each longitude in the
218 standard manner (Wilks 2006, eq. 7.34). The Brier Skill Score (BSS; ibid, eq. 7.35)
219 was computed from the sums of Brier Scores of the forecast and climatology over

220 each longitude within a sector. Forecast probabilities of blocking were set directly
221 by ensemble relative frequency. For example, if 3 of the 11 members were
222 diagnosed as having blocked conditions at a particular longitude, the forecast
223 probability was set to 3/11. Since climatological blocking frequency did not vary
224 radically within each sector (see Fig. 2 below), there was little reason to apply more
225 complex procedures (e.g., Hamill and Juras 2006) to account for variations in
226 climatological event frequency for each sector when calculating the BSS. Skill will
227 also be evaluated for block onset and cessation for long-lived blocks. The subset of
228 dates classified as “block onset” for long-lived blocks were those dates where,
229 within a given sector, there were < 20 degrees of longitude that were blocked the
230 day before but \geq 20 degrees blocked on that day, and where \geq 20 degrees continued
231 to be blocked for at least the subsequent 9 days; this method is similar to the
232 method of Colucci and Alberta (1996), though they require the duration to be only 5
233 days. Dates classified as “block cessation” were those dates where, for the first date
234 after a defined block onset, < 20 degrees of longitude were blocked. Finally, we will
235 also evaluate the BSS of a “perfect-model” forecast (Buizza 1997). One of the 11
236 forecast members replaced the analyzed state, and then the existence / non-
237 existence of a block was computed from this replacement data. Probabilities were
238 estimated from the diagnoses of blocking from the remaining 10 forecast members.
239 Reliability diagrams (Wilks 2006, p. 287) for blocking probabilities were also
240 calculated in the standard manner.
241 The MJO was also evaluated with a variety of metrics. Following now
242 standard metrics defined by Lin et al. (2008), the correlation skill (COR) and root-

243 mean-square error (RMSE) of the RMMs were calculated over all ensemble
 244 members as:

$$245 \quad COR(\tau) = \frac{\sum_{j=1}^{11} \sum_{i=1}^N [RMM1_i^a RMM1_{ij}^f + RMM2_i^a RMM2_{ij}^f]}{\sqrt{\sum_{j=1}^{11} \sum_{i=1}^N [RMM1_i^a]^2 + RMM2_i^a]^2} \sqrt{\sum_{j=1}^{11} \sum_{i=1}^N [RMM1_{ij}^f]^2 + [RMM2_{ij}^f]^2}} \quad (1)$$

246 and

$$247 \quad RMSE(\tau) = \sqrt{\frac{1}{Nx11} \sum_{j=1}^{11} \sum_{i=1}^N \left\{ [RMM1_i^a - RMM1_{ij}^f]^2 + [RMM2_i^a - RMM2_{ij}^f]^2 \right\}} \quad (2)$$

248 where τ is the lead time in days, $RMM1_i^a$ and $RMM2_i^a$ are the analyzed RMM1 and
 249 RMM2 projections for the i th of N sample days. $RMM1_{ij}^f$ and $RMM2_{ij}^f$ are their
 250 respective, time-concurrent forecast projections for each of the j th of 11 members.
 251 The COR and $RMSE$ will also be calculated for “high-amplitude” and “low-amplitude”
 252 events. The i th forecast sample is evaluated as high amplitude or low amplitude
 253 depending on the magnitude of the associated analysis. The sample is “large
 254 amplitude” if $\sqrt{[RMM1_i^a + RMM2_i^a]^2} \geq 1.0$ and “small amplitude” (or possibly no
 255 MJO) if < 1.0 .

256 We were also interested in the ability of the ensemble to provide high-quality,
 257 reliable probabilistic guidance. For the MJO, reliability (or more accurately,
 258 “consistency”) was evaluated with rank histograms (Hamill 2001). As the rank
 259 histograms for RMM1 and RMM2 were similar, the average of the two was reported.
 260 Rank histograms were also generated for MJO phase and amplitude. In generating
 261 the rank histograms, the observed value of RMM1 and RMM2 were assumed to be
 262 perfect, so no noise was introduced to ensemble members to potentially account for
 263 inaccuracies in the analyses of RMMs (ibid, Fig. 6 therein). The phase and

264 amplitude propagation characteristics of the reforecasts will also be examined; the
265 specific methodology for these will be discussed at the relevant point in the results
266 section.

267 A continuous ranked probability skill score (CRPSS; Wilks 2006, p. 302) was
268 also calculated for MJO forecasts against two reference forecasts, climatology and
269 lagged persistence. As with rank histograms, continuous ranked probability score
270 (CRPS) statistics for the forecast and the reference were calculated separately for
271 RMM1 and RMM2 and then summed. Let Φ_i^f represent the ensemble forecast
272 cumulative distribution function (CDF) for RMM1 or RMM2 for the ith of N samples.
273 Similarly, let Φ_i^a represent the analyzed CDF, which is a step function, 0 below the
274 RMM value and 1 above. Φ_i^r is the CDF of the reference forecast, be it the
275 unconditional climatology for Dec-Jan-Feb or a lagged persistence forecast. Then
276 the CRPSS was calculated as

$$277 \quad CRPSS = 1.0 - \frac{CRPS^f}{CRPS^r}, \quad (3)$$

278 where the $CRPS^f$ was the CRPS of the forecast and $CRPS^r$ was the CRPS of the
279 reference. $CRPS^f$ was calculated via

$$280 \quad CRPS^f = \frac{1}{N} \sum_{i=1}^N \int_{RMM=-\infty}^{\infty} (\Phi_i^f - \Phi_i^a)^2 dRMM, \quad (4)$$

281 with the $CRPS^r$ calculated similarly. For the climatological reference forecast, a
282 Gaussian distribution was fitted to the climatology individually for RMM1 and
283 RMM2, which was then used to generate the CDFs. The lagged persistence was
284 based directly on the “PCRLAG” approach documented in Seo et al. (2009, their eq.

285 1). Future RMM values were predicted using a multiple linear regression model
286 using the current RMM values and their values 5, 10, 15, 20 and 25 days in the past.

287 Some figures below will include PDFs of the daily change in the angle θ and
288 magnitude of RMM in the (RMM1, RMM2) phase space. θ measures the rotation
289 from the positive RMM1 axis, and is defined as

290
$$\theta = \tan^{-1}(RMM2/RMM1), \quad (5)$$

291 which can be defined uniquely in the interval $-180^\circ \leq \theta < 180^\circ$ with knowledge of
292 the signs of RMM2 and RMM1. Angles of -180° to -135° correspond to “phase 1” in
293 the (RMM1, RMM2) phase space (see Wheeler and Hendon 2004, Fig. 7); angles of -
294 135° to -90° correspond to “phase 2,” and so on. The PDFs of θ and RMM magnitude
295 were estimated with kernel density estimation using a Cressman-shaped kernel
296 (Cressman 1959) that tapered to 0.0 at 2° or 0.1 units of RMM magnitude.

297

298 **3. Results.**

299 *a. Blocking forecasts.*

300 Figure 2a shows the blocking frequency in the GEFS reforecasts for selected
301 forecast lead times. Overall, the GEFS replicated blocking frequency reasonably
302 accurately, though for lead times of +6 days and beyond it under-forecasted
303 blocking frequency in the Euro-Atlantic sector by up to 25%. While the blocking
304 frequency curves are relatively smooth over the multi-decadal period, this disguises
305 tremendous inter-annual variability. Figure 2b shows the yearly blocking
306 frequencies, spatially smoothed slightly to aid in interpretability. For a given
307 longitude, blocking frequencies can vary by an order of magnitude or more from one

308 year to the next. Figure 3 shows that overall positive blocking forecast skill was
309 retained through day +13, but this skill was far short of the skill that was possible
310 under-perfect model assumptions. For example, the perfect-model skill at day +7
311 was as large as the actual skill at ~ day +3.5. It is somewhat likely that the perfect-
312 model estimate of forecast skill is somewhat too large, too, due to the tendency for
313 the ensemble forecasts to cluster together, such that their spread is not statistically
314 consistent with their ensemble-mean error (e.g., Bougeault et al. 2010, Figs. 2-3).
315 Figure 3 also shows that blocking onset and cessation were somewhat less well
316 forecast than the overall forecasts of blocking. The skill curves for onset and
317 cessation were noisier because of the greatly reduced sample size, even with 28
318 winter seasons of data. Figure 4 shows that there was a substantial amount of
319 variability in the skill of blocking when the data was sorted by half-decadal periods,
320 plus 2010-2012. In the Pacific sector, the blocking skill in the most recent 3 years
321 was the largest, but the 2005-2009 period was intermediate in skill and actually
322 comparable to the skill during the 1985-1989 period. However, in the Euro-
323 Atlantic sector, the skill for the 1985-1989 period was substantially smaller than for
324 the subsequent half decades. The dashed lines in Fig. 4 provide the half-decadal skill
325 results under the perfect-model assumption. These show some natural variability in
326 skill at half-decadal timescales. Note that the 1985-1989 period, for example, had
327 the lowest perfect-model predictability in the Euro-Atlantic sector, which probably
328 contributed to its especially low real-model skill, while the 2010-2012 period had
329 among the highest perfect-model skill in the Pacific, indicating that the actual high
330 forecast skill for this period was in part natural variability. The perfect-model

331 results here also suggest that even in the best of circumstances, blocking forecast
332 skill as defined here is limited to approximately two weeks.

333 Figure 5 presents reliability diagrams for the blocking forecasts at various
334 lead times. At the earlier lead times the blocking forecasts are mostly reliable, but
335 the reliability decreases so that by day +15 the forecasts are rather unreliable. As
336 expected, blocking forecast sharpness decreases over time, as seen in the usage
337 frequency histograms. There are probably many reasons for the lack of reliability,
338 including all the usual suspects with unreliable ensemble forecasts; the moderate
339 resolution of the forecast model, the deficiencies in parameterizations, including
340 sometimes inappropriate deterministic formulations (Palmer 2012), the sub-
341 optimal initialization of both the control and perturbations in the ensemble system,
342 and, as we shall see below, faulty representations of interactions with the MJO.

343

344 *b. MJO forecasts.*

345 Figure 6 shows the evolution from both analyzed and deterministic forecasts
346 where the vector (RMM1, RMM2) of initial conditions was within 0.5 units of (1.5, -
347 1.5), i.e., within the purple circle on the figure. The samples are all various dates
348 from Dec-Jan-Feb 1985-2012, and samples that were less than 5 days apart from
349 another sample were eliminated. It appears that the collection of forecasts
350 propagate somewhat more regularly than the collection of analyzed states, and
351 perhaps the forecasts lose some amplitude. Figures 7 and 8 attempt to quantify this,
352 providing a PDF of the daily change in overall RMM angle and magnitude,
353 respectively, as well as the change attributable specifically to wind and OLR

354 components of the RMM. The cases used to populate Figs. 7 and 8 were selected in
355 an attempt to isolate situations where there was a real MJO between Africa and the
356 Maritime continent, and where it had robust associated convection. First, we note
357 that a given RMM vector can be decomposed into a component due to the OLR and a
358 component due to the winds. Consequently, a subset of Dec-Jan-Feb 1985-2012
359 cases was selected that met the following criteria: (a) an overall RMM amplitude of
360 greater than 1.0, (b) an $RMM2 < 0.0$, and (c) an amplitude of the OLR component $>$
361 0.5. There were 90 such dates. Consider the overall phase change in Fig. 7a; after a
362 reasonable mean phase change the first day, for subsequent leads the mean phase
363 change was substantially smaller for the forecast than for the analyzed, with a phase
364 change of $5\text{-}10^\circ$ per day for the analyzed but $3\text{-}6^\circ$ for the forecast. Forecast MJOs
365 propagated too slowly, on average, a problem noted also in the CFS version 2 (CFS
366 v2; Fu et al. 2013). The forecast propagation of the OLR component of RMM was
367 slower than the wind component, though this was true to a lesser extent with
368 analyzed data as well. This slower propagation can also be diagnosed, for example,
369 from the reanalyzed rainfall and wind lag correlations shown in Fig. 3 from Weaver
370 et al. (2011). The overall RMM phase change distribution also more closely
371 resembles the wind change than the OLR change, evidence of the domination of
372 RMM calculation by the wind component (Straub 2013). By inspection, the
373 distributions of phase changes of the forecast PDFs were not dramatically narrower
374 than of the analyzed PDFs; the main deficiency was a biased mean, not a lack of
375 spread. Considering the overall magnitude change in Fig. 8a, both forecast and

376 analyzed exhibit a similar small decrease in magnitude, though the PDF for the
377 forecast appears slightly more narrow and peaked than for the analyzed.

378 Figure 9 provides more evidence of the deficiencies of phase and amplitude
379 in the MJO forecasts. Here, for panels (a) and (b), the set of dates in Dec-Jan-Feb
380 1985-2012 was identified that had an (RMM1, RMM2) in phase 1, i.e., $-180^\circ \leq \theta \leq -$
381 135° . Panels (c) and (d) use initial dates with an (RMM1, RMM2) in phase 4, i.e., $-$
382 $45^\circ \leq \theta \leq -0^\circ$. The panels then present the lagged OLR and 850 hPa wind
383 component anomalies during the subsequent 15 days, using analyzed data (panels
384 (a) and (c)) and forecast data (panels (b) and (d)). For a comparison to CFS v2, see
385 Fig. 9 of Wang et al. (2013). Consider first the analyzed data for phase 1 in Fig. 9(a).
386 Here, a small initial MJO convective anomaly was roughly centered over Africa. 15
387 days later a stronger cold anomaly had developed that was centered more over the
388 western Indian Ocean. Low-level zonal wind convergence anomalies (as diagnosed
389 from the 0.0 wind contour) moved from $\sim 30^\circ$ E at the initial time to $\sim 100^\circ$ E 15
390 days later. Considering the associated forecast data in Fig. 9(b), the initial cold OLR
391 anomaly and the wind anomaly did not grow during the subsequent 15 days, nor did
392 the anomalies propagate as rapidly; the forecast model generally misses the
393 initiation of the MJO, and in this regard performs worse than CFS v2 (*ibid*). Fig. 9(c)
394 presents the analyzed anomaly data for phase 4, when the center of the convection
395 associated with MJO is approaching the maritime continent. The maintenance and
396 propagation of cold OLR anomalies can be clearly seen, as well as the gradual shift of
397 low-level convergence toward the international dateline during the subsequent 15
398 days. Considering the associated forecast data (Fig. 9(d)), here the MJO-associated

399 convection and the strength of forecast wind anomalies were better maintained
400 relative to phase 1 forecasts. However, the forecast wind anomalies still propagated
401 too slowly, as diagnosed from the movement of the 0 anomaly contour to only ~
402 150° E during the subsequent 15 days.

403 Not only did the forecast MJO wind and precipitation features propagate too
404 slowly, but also the precipitation forecasts exhibited significant unconditional bias
405 (Fig. 10). Precipitation amounts were dramatically over-forecast at the early
406 forecast leads. The general pattern of the daily GPCP precipitation amount
407 climatology was reasonably replicated in day +0 to 1 forecast (Fig. 10b), but the
408 average daily forecast precipitation amounts was commonly > 50% too large. By
409 the beginning of the second week of the forecast (Fig. 10c), the over-forecast bias
410 was reduced, but there was less resemblance with the analyzed precipitation
411 pattern. For example, the connection of the South Pacific Convergence Zone (SPCZ;
412 Folland et al. 2002) to the inter-tropical convergence zone was missing, and the
413 forecast SPCZ was unduly zonally oriented, as it often is in climate simulations
414 (Brown et al. 2011). These pattern changes and an excess of forecast precipitation
415 in the central to eastern Pacific were apparent at the end of week +2 (Fig. 10d).

416 An additional deficiency of the probabilistic MJO forecasts was their under-
417 dispersion and/or conditional bias. This can be seen by examining the rank
418 histograms from the ensemble predictions of MJO RMMs (Fig. 11). All of the rank
419 histograms were U-shaped, which was most pronounced for the short-lead RMM
420 amplitude forecasts. The rank histograms indicate that there was unrealistic
421 consistency of magnitudes among the forecasts; the ensemble prediction system did

422 not adequately simulate the forecast processes that contribute to diversity in MJO
423 magnitudes.

424 Despite the significant biases, the forecasts still exhibit skill in the first week.
425 Figure 12 presents the CRPSS of the forecasts measured relative to an unconditional
426 climatology and relative to the regression-based lagged persistence model. The
427 lagged persistence model presented a tougher reference standard, so forecasts
428 exhibited less skill in comparison to this. Skill diminished to near 0 by day +11 with
429 respect to lagged persistence and by day +14 with respect to climatology. While
430 lagged persistence represented a tougher reference, we note that the first-
431 generation Climate Forecast System at NCEP (Wang et al. 2005, Saha et al. 2006)
432 produced forecasts that had higher errors and less correlation skill than the lagged
433 persistence at all forecast leads. Thus, the current GEFS provides substantial
434 improvement in the simulation of the MJO relative to the first-generation CFS (CFS
435 v1).

436 Consider now the correlation skill and RMSE of MJO forecasts (Fig. 13).
437 Overall correlation skill and RMSE were comparable to those from the more
438 accurate models shown in Matsueda and Endo (2011), though not as large as for the
439 CFS v2 (Wang et al. 2013, their Fig. 3). Perhaps this difference is in part due to the
440 use of a coupled ocean and atmosphere in the CFS v2, unlike in the GEFS used here.
441 There was more correlation skill in the wind components of the RMM than in the
442 OLR component, as shown in Fig 13(a). Note that the Wheeler-Hendon RMM index is
443 dominated by its wind component (Straub 2013). The greater skill for wind likely
444 relates to the better ability of models to maintain and evolve the rotational

445 component of the wind over the globe, perhaps due to improved initial conditions.
446 The correlation skill was much lower for the first two half-decadal periods, 1985-
447 1989 and 1990-1994, than for the subsequent periods, with the exception of 2010-
448 2012, though the RMS error was not larger for the first decade of the forecast (Figs.
449 13 (c) and (d)). The generally greater skill since 2000 is likely due in part to
450 assimilating a greater number and variety of satellite observations, especially
451 radiance data from polar-orbiting satellites (Gelaro et al. 2010). Figures 13 (e) and
452 (f) show that forecasts initialized from analyses with high amplitude of RMM
453 exhibited more skill but also higher RMSE than lower-amplitude forecasts. This was
454 also shown in Lin et al. (2008, their Fig. 13) for the Canadian models.

455

456 *c. Interactions between blocking and the MJO.*

457 Finally, we briefly consider the ability of the forecast model to successfully
458 replicate the ability to discern changes in blocking frequency for different phases of
459 strong MJOs. Strong MJOs are defined as the set of dates where the magnitude of
460 the RMM is in the upper quartile of its distribution, forecast or observed. Figure 14
461 presents the results. Consider panel (a). Here the change in blocking frequency for
462 a strong MJO relative to the unconditional blocking frequency is shown as a function
463 of longitude (abscissa) and of the analyzed phase θ (ordinate) of the MJO, as defined
464 in section 2c. To provide an adequate sample size for a given θ , the data plotted for
465 a given θ actually includes analyzed samples with similarly diagnosed θ , specifically
466 where $-22.5^\circ \leq \theta \leq 22.5^\circ$. Note some interesting characteristics in the analyzed
467 relationship of blocking frequency changes. As θ varies between 0° and 120° (i.e.,

468 the MJO's center moves from the Maritime continent to western N. America), at 0°
469 longitude, the blocking frequency changes from a strongly negative anomaly in
470 blocking frequency to a strongly positive anomaly. Restated, analyzed Euro-Atlantic
471 blocking frequency changes from below its long-term average to above average as
472 the MJO moves east from the Maritime continent. We note that this is consistent
473 with previous results, such as the 500 hPa anomaly composite for various phases of
474 the MJO in Lin et al. (2009, their Fig. 4). Our Fig. 14(b)-(d) then show the respective
475 blocking frequency anomalies when the *forecast* MJO phase is of the noted angle,
476 and when the *forecast* MJO magnitude is greater than the upper quartile of the
477 forecast distribution. The day +4 forecast in panel (b) still replicates many of the
478 essential anomalies of the analyzed, including the shift from a strongly positive
479 blocking frequency anomaly to a strongly negative anomaly along the Greenwich
480 meridian as θ varies between 0 and 120 degrees. In some regards, this may not
481 fully represent a particular ability of the forecast model to predict the interaction,
482 but may also be an artifact of the persistent nature of blocks. Much of the frequency
483 anomaly detail is lost by day +8, and the day +16 forecasts show no apparent
484 relation to the analyzed. From this, we can conclude that the internal dynamics of
485 the GEFS do not represent very well the processes that lead to inter-relationships
486 between blocking and the MJO.

487

488 **4. Discussion and conclusions**

489 Some modes of atmospheric variability are uncommon enough and/or
490 operate on long-enough timescales that a short time series of past forecasts will not

491 prove sufficient for diagnosing their characteristics. Atmospheric blocking and the
492 MJO are two such phenomena. In this paper we have shown how a very long time
493 series of ensemble forecast guidance facilitates a greater understanding of the
494 forecastability and predictability of these phenomena. In this case, the long time
495 series was provided by a 28-year data set of reforecasts from the NCEP Global
496 Ensemble Forecast System. The paper more specifically explored Northern
497 Hemispheric blocking, the MJO, and their interaction during Dec-Jan-Feb 1985-2012
498 period.

499 With regards to blocking, the reforecasts showed that the GEFS slightly
500 under-forecasted blocking frequency at longer leads in the Euro-Atlantic sector.
501 Furthermore, the inter-annual variability of blocking frequency was shown to be
502 quite large, demonstrating how difficult it can be to achieve a representative sample
503 with only a few years of data. The predictive skill of the probabilistic forecasts of
504 actual blocking was substantially smaller than its perfect-model skill, whereby a
505 member of the ensemble was used as a synthetic verification. This indicates that
506 there is still tremendous potential for improvement in blocking forecasts. However,
507 it is also likely that the perfect-model results present a somewhat over-optimistic
508 estimate of the upper range of forecast skill. The GEFS system and most other
509 ensemble systems are under-dispersive, and as such, the members of the ensemble
510 unduly resemble each other, inflating the perfect-model skill estimates. It was also
511 found that block onset and cessation were forecast somewhat less well than block
512 maintenance, and there was substantial variability of blocking skill between half-
513 decadal periods. Finally, the reliability of probabilistic blocking forecasts degraded

514 with increasing lead time, and as expected, blocking forecasts became progressively
515 less sharp, i.e., forecast probabilities were less often 0.0 and 1.0 and more often
516 resembled the model climatology.

517 Forecasts of strong MJOs propagated too slowly, especially the component
518 associated with outgoing longwave radiation (OLR), i.e., convection. Deep tropical
519 convection appeared to have other systematic biases in the GEFS; in general, there
520 was too much tropical precipitation forecast in the Indian and Pacific Oceans,
521 especially for the shorter forecast leads. The ensemble predictions were biased
522 and/or under-dispersive, manifested in U-shaped rank histograms of MJO indices.
523 Forecasts of the magnitude of the MJO's leading EOFs were especially U-shaped. Bi-
524 variate MJO correlation skill was found to be larger for the wind component than for
525 the OLR component, and skill was larger for the higher-amplitude MJO events. Skill
526 varied significantly between half-decadal periods, with the period 1985-1994 and
527 2010-2012 exhibiting lower MJO skill than the 1995-2009 period. Probabilistic
528 skill of the MJO forecast was modest, and skill was larger when measured relative to
529 climatology than when measured relative to a lagged persistence forecast. Finally,
530 for longer-lead forecasts, the GEFS demonstrated little ability to replicate the
531 changes in blocking frequency due to a strong MJO that were noted in analyzed data.

532 This paper has discussed forecast skill without providing analysis of the
533 potential forecast systematic errors that may lead to deficiencies in blocking, the
534 MJO, and their inter-relationships. This much more challenging work is left as future
535 research. We do hope that we have laid out a first step, demonstrating the

536 predictability and forecastability of these phenomena using the newly created GEFS
537 reforecast data set.

538

539 **Acknowledgments:**

540 The reforecast data set used here was created under a 2010 Department of Energy
541 Advanced Leadership Computing Challenge supercomputer grant. This manuscript
542 was partly supported by a NOAA USWRP grant on methods for estimating model
543 uncertainty. We thank John Cortinas, the NOAA program manager, for his assistance
544 and guidance on USWRP funding.

545

546 **References:**

- 547 Barnes, E., A., J. Slingo, and T. Woolings, 2012: A methodology for the comparison of
548 blocking climatologies across indices, models, and climate scenarios. *Clim.
549 Dyn.*, **38**, 2467-2481. DOI: 10.1007/s00382-011-1243-6.
- 550 Bougeault, P., and others, 2010: The THORPEX interactive grand global ensemble.
551 *Bull. Amer. Meteor. Soc.*, **91**, 1059-1072, doi: 10.1175/2010BAMS2853.1.
- 552 Brown, J. R., S. B. Power, F. P. Delage, R. A. Colman, A. F. Moise, and B. F. Murphy,
553 2011. Evaluation of the South Pacific Convergence Zone in IPCC AR4 climate
554 model simulations of the Twentieth Century. *J. Climate*, **24**, 1565-1582.
- 555 Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill
556 distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.*,
557 **125**, 99-119.
- 558 Cassou, C., 2008: Intraseasonal interaction between the Madden-Julian Oscillation
559 and the North Atlantic Oscillation. *Nature*, **455**, 523-527.
- 560 Colucci, S. J., and T. L. Alberta, 1996: Planetary-scale climatology of explosive
561 cyclogenesis and blocking. *Mon. Wea. Rev.*, **124**, 2509-2520.
- 562 Crueger, T. , B. Stevens, R. Brokopf, 2013: The Madden-Julian Oscillation in ECHAM6
563 and the introduction of an objective MJO metric. *J. Climate*, in press.
564 Available at [http://journals.ametsoc.org/doi/pdf/10.1175/JCLI-D-12-
565 00413.1](http://journals.ametsoc.org/doi/pdf/10.1175/JCLI-D-12-00413.1)
- 566 Drosdowsky, W., and L. E. Chambers, 2001: Near-global sea surface temperature
567 anomalies as predictors of Australian seasonal rainfall. *J. Climate*, **14**, 1677-
568 1687.

- 569 Ferranti, L., Palmer, T. N., Molteni, F., and Klinker, E., 1990: Tropical-extratropical
570 interaction associated with the 30-60 day oscillation and its impact on
571 medium and extended-range prediction. *J. Atmos. Sci.*, **125**, 2177-2199.
- 572 Folland, C. K., J. A. Renwick, M. J. Salinger, and A. B. Mullan, 2002: Relative influences
573 of the interdecadal Pacific Oscillation and ENSO in the South Pacific
574 Convergence Zone. *Geophys. Res. Letters*, **29**, 21-1-21-4. Doi:
575 10.1029/2011GL014201.
- 576 Fu, X., J.-Y. Lee, P.-C. Hsu, H. Taniguchi, B. Wang, W. Wang, and S. Weaver, 2013:
577 Multi-model MJO forecasting during DYNAMO/CINDY period. *Clim. Dyn.*, **41**,
578 1067-1081.
- 579 Gelaro, R., R. H. Langland, S. Pellerin, and R. Todling, 2010: The THORPEX
580 observation impact Intercomparison experiment. *Mon. Wea. Rev.*, **138**, 4009-
581 4025.
- 582 Gottschalck, J., M. Wheeler, K. Weickmann, F. Vitart, N. Savage, H. Lin, H. Hendon, D.
583 Waliser, K. Sperber, M. Nakagawa, C. Prestrelo, M. Flatau, and W. Higgins,
584 2010: A framework for assessing operational Madden-Julian Oscillation
585 forecasts. *Bull. Amer. Meteor. Soc.*, **91**, 1247-1258.
- 586 Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble
587 forecasts. *Mon. Wea. Rev.*, **129**, 550-560.
- 588 Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: is it real skill or is it the
589 varying climatology? *Quart. J. Royal Meteor. Soc.*, **132**, 2905-2923.

- 590 Hamill, T. M., J. S. Whitaker, D. T. Kleist, M. Fiorino, and S. G. Benjamin, 2011:
591 Predictions of 2010's tropical cyclones using the GFS and ensemble-based
592 data assimilation methods. *Mon. Wea. Rev.*, **139**, 3243-3247.
- 593 Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Jr., Y.
594 Zhu, and W. Lapenta, 2013: NOAA's second-generation global medium-range
595 ensemble reforecast data set. *Bull. Amer. Meteor. Soc.*, **94**, in press.
596 DOI:10.1175/BAMS-D-12-00014.1
- 597 Hendon, H. H., B. Liebmann, M. Newman, J. D. Glick, 2000: Medium-range forecast
598 errors associated with active episodes of the Madden-Julian Oscillation. *Mon.*
599 *Wea. Rev.*, **128**, 69-86.
- 600 Huffman, G.J., R.F. Adler, M. Morrissey, D.T. Bolvin, S. Curtis, R. Joyce, B McGavock, J.
601 Susskind, 2001: Global precipitation at one-degree daily resolution from
602 multi-satellite observations. *J. Hydrometeor.*, **2**, 36-50.
- 603 Jia, X., C. Li, N. Zhou, and J. Ling, 2010: The MJO in an AGCM with three different
604 cumulus parameterization schemes. *Dyn. of Atmos. Oceans*, **49**, 141-163.
- 605 Jones, C., D. E. Waliser, K. M. Lau, and W. Stern, 2004: The Madden-Julian Oscillation
606 and its impact on Northern Hemisphere weather predictability. *Mon. Wea.*
607 *Rev.*, **132**, 1462-1471.
- 608 Jung, T., M. J. Miller, T. N. Palmer, P. Towers, N. Wedi, D. Achuthavarier, J. M. Adams,
609 B. A. Cash, J. L. Kinter III, L. Marx, C. Stan, and K. I. Hodges, 2012: High-
610 resolution global climate simulations with the ECMWF model in project
611 Athena: experimental design, model climate, and seasonal forecast skill. *J.*
612 *Climate*, **25**, 3155-3172. DOI: 10.1175/JCLI-D-11-00265.1

- 613 Kalnay, E., and co-authors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull.*
614 *Amer. Meteor. Soc.*, **77**, 437-471.
- 615 Kang, I.-S., and H.-M. Kim, 2010: Assessment of MJO predictability for Boreal winter
616 with various statistical and dynamical models. *J. Climate*, **23**, 2368-2378.
- 617 Kiladis, G. N., and K. M. Weickmann, 1992: Circulation anomalies associated with
618 tropical convection during norther winter. *Mon. Wea. Rev.*, **120**, 1900-1923.
- 619 Kim, D., K. Sperber, W. Stern, D. Waliser, I. S. Kang, E. Maloney, W. Wang, K.
620 Weickmann, J. Benedict, M. Khairoutdinov, M.-I. Lee, R. Neale, M. Suarez, K.
- 621 Thayer-Calder, and G. Zhang, 2009: Application of MJO simulation
622 diagnostics to climate models. *J. Climate*, **22**, 6413-6436.
- 623 Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009:
624 Introduction of the GSI into the NCEP global data assimilation system. *Wea.*
625 *Forecasting*, **24**, 1691-1705.
- 626 Knutson, T. R. and K. M. Weickmann, 1987: 30-60 day atmospheric oscillations:
627 Composite life cycles of convection and circulation anomalies. *Mon. Wea. Rev.*,
628 **115**, 1407-1436.
- 629 Liebmann, B., and D. L. Hartmann, 1984: An observational study of tropical-
630 midlatitude interaction on intraseasonal timescales during winter. *J. Atmos.*
631 *Sci.*, **41**, 3333-3350.
- 632 L'Heureux, M. L., and R. W. Higgins, 2008: Boreal winter links between the Madden-
633 Julian Oscillation and the Arctic Oscillation. *J. Climate*, **21**, 3040-3050.

- 634 Lin, H., G. Brunet, and J. Derome, 2008: Forecast skill of the Madden-Julian
635 Oscillation in two Canadian atmospheric models. *Mon. Wea. Rev.*, **136**, 4130-
636 4149.
- 637 Lin, H., G. Brunet, and J. Derome, 2009: An observed connection between the North
638 Atlantic Oscillation and the Madden-Julian Oscillation. *J. Climate*, **22**, 364-380.
- 639 Madden, R. L., and Julian, P. R., 1971: Detection of a 40-50 day oscillation in the zonal
640 wind in the tropical Pacific. *J. Atmos. Sci.*, **5**, 702-708.
- 641 Maloney, E. D., and Hartmann, D. L., 2000a: Modulation of the eastern North Pacific
642 hurricanes by the Madden-Julian Oscillation. *J. Climate*, **8**, 1757-1774.
- 643 Maloney, E. D., and Hartmann, D. L., 2000b: Modulation of hurricane activity in the
644 Gulf of Mexico by the Madden-Julian Oscillation. *Science*, **287**, 2002-2004.
- 645 Matsueda, M., and Endo, H., 2011: Verification of medium-range MJO forecasts with
646 TIGGE. *Geophys. Res. Letters*, **38**, 1-6. Doi:10.1029/2011GL047480.
- 647 Mauritsen, T., and E. Källén, 2004: Blocking prediction in an ensemble forecasting
648 system. *Tellus*, **56A**, 218-228.
- 649 Palmer, T. N., 2012: Towards the probabilistic earth-system simulator: a vision for
650 the future of climate and weather prediction. *Quart. J. Royal Meteor. Soc.*, **138**,
651 841-861. DOI: 10.1002/qj.1923.
- 652 Pelly, J. L., and B. J. Hoskins, 2003a: How well does the ECMWF ensemble prediction
653 system predict blocking? *Quart. J. Royal Meteor. Soc.*, **129**, 1683-1702.
- 654 Pelly, J. L., and B. J. Hoskins, 2003b: A new perspective on blocking. *J. Atmos. Sci.*, **60**,
655 743-755.

- 656 Riddle, E. E., M. B. Stoner, N. C. Johnson, M. L. L'Heureux, D. C. Collins, and S. B.
657 Feldstein, 2013: The impact of the MJO on clusters of wintertime circulation
658 anomalies over the North American Region. *Clim. Dyn.*, **40**, 1749-1766, DOI:
659 10.1007/s00382-102-1493-y.
- 660 Saha, S., and co-authors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**,
661 3483-3517.
- 662 Saha, S., and co-authors, 2010: The NCEP climate forecast system reanalysis. *Bull.*
663 *Amer. Meteor. Soc.*, **91**, 1015-1057.
- 664 Seo, K.-H., W. Wang, J. Gottschalck, Q. Zhang, J.-K. Schemm, W. R. Higgins, and A.
665 Kumar, 2009: Evaluation of MJO forecast skill from several statistical and
666 dynamical forecast models. *J. Climate*, **22**, 2372-2388.
- 667 Straub, K. H., 2013: MJO initiation in the realtime multivariate MJO index. *J. Climate*,
668 **26**, 1130-1151.
- 669 Tibaldi, S., and F. Molteni, 1990: On the operational predictability of blocking.
670 *Tellus*, **42A**, 343-365.
- 671 Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.*, **78**, 2771-
672 2777.
- 673 Vitart, F., and F. Molteni, 2010: Simulation of the Madden-Julian Oscillation and its
674 teleconnections in the ECMWF forecast system. *Quart. J. Royal Meteor. Soc.*,
675 **136**, 842-855.
- 676 Waliser, D., and others, 2009: MJO simulation diagnostics. *J. Climate*, **22**, 3006-
677 3030.

- 678 Wang, W., S. Saha, H.-L. Pan, S. Nadiga, and G. White, 2005: Simulation of ENSO in the
679 new NCEP Coupled Forecast System Model (CFS03). *Mon. Wea. Rev.*, **133**,
680 1574-1593.
- 681 Wang, W., M.-P. Hung, S. J. Weaver, A. Kumar, and X. Fu, 2013: MJO prediction in the
682 NCEP Climate Forecast System version 2. *Clim. Dyn.* DOI: 10.1007/s00382-
683 013-1806-9.
- 684 Watson, J. S., and S. J. Colucci, 2002: Evaluation of ensemble predictions of blocking
685 in the NCEP global spectral model. *Mon. Wea. Rev.*, **130**, 3008-3021.
- 686 Weaver, S. J., W. Wang, M. Chen, and A. Kumar, 2011: Representation of MJO
687 variability in the NCEP climate forecast system. *J. Climate*, **24**, 4676-4694.
- 688 Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the
689 ensemble transform (ET) technique in the NCEP global operational forecast
690 system. *Tellus A*, **60**, 62-79.
- 691 Weickmann, K., and E. Berry, 2009: The tropical Madden-Julian Oscillation and the
692 global wind oscillation. *Mon. Wea. Rev.*, **137**, 1601-1614.
693 DOI: <http://dx.doi.org/10.1175/2008MWR2686.1>
- 694 Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO
695 index: development of an index for monitoring and prediction. *Mon. Wea.
696 Rev.*, **132**, 1917-1932.
- 697 Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences* (2nd Ed.).
698 Academic Press, 627 pp.
- 699 Yasunari, T., 1979: Cloudiness fluctuations associated with the Northern
700 Hemisphere summer monsoon. *J. Meteor. Soc. Japan*, **58**, 225-229.

701 **FIGURE CAPTIONS**

702

703 **Figure 1:** Composite of Northern Hemisphere 500 hPa geopotential height patterns

704 under (a) blocked flow at 180° E longitude, and (b) unblocked flow at 180° E

705 longitude.

706 **Figure 2:** (a) Blocking frequency as determined from analyses and reforecasts as a

707 function of forecast lead time. Areas shaded in gray denote the two sectors in

708 subsequent figures, the Pacific and Euro-Atlantic sectors. (b) Analyzed, spatially

709 smoothed yearly blocking frequencies for each year between 1985 and 2012.

710 **Figure 3:** Brier Skill Scores of blocking probability forecasts for (a) Pacific, and (b)

711 Euro-Atlantic sectors.

712 **Figure 4:** As in Fig. 3, but for Brier Skill Scores of blocking probability forecasts by

713 half decade for (a) Pacific, and (b) Euro-Atlantic sectors. Solid lines present the skill

714 scores for the actual reforecasts, dashed lines present the skill under perfect-model

715 assumptions.

716 **Figure 5:** Reliability diagrams for blocking probability forecasts for (a) +3 day

717 forecast, (b) +6 day forecast, (c) +9 day forecast, (d) +12 day forecast, and (e) +15

718 day forecast. Dotted red line denotes the skill in the Euro-Atlantic sector and dotted

719 blue line denotes the skill in the Pacific sector. Red and blue bars indicate the

720 frequency of usage of each forecast probability category for the Euro-Atlantic and

721 Pacific sectors, respectively.

722 **Figure 6:** RMM1 and RMM2 phase plots for (a) analyzed, and (b) control forecasts

723 whose initial states are within the purple circle. Differences in time from the initial

724 time conveyed by the colors of lines and dots, with the legend indicating the lead
725 time in days. Phases 1 to 8 are marked in the corners of the diagram.

726 **Figure 7:** PDF of daily change in angle of RMM vector ($\Delta\theta$), measured in degrees, for
727 (a) overall RMM, (b) OLR contribution to RMM, and (c) wind-component
728 contribution to RMM. Only Dec-Jan-Feb 1985-2012 dates which had an initial RMM
729 magnitude of greater than 1.0, an RMM2 component < 0.0 , and an initial OLR
730 component of > 0.5 were included as samples. Dots indicate the mean of the PDF for
731 a given day, horizontally offset slightly so that dots do not overlap.

732 **Figure 8:** As in Fig. 7, but for the PDF of the change in magnitude of RMM.

733 **Figure 9:** Composite of lagged filtered OLR anomalies (shaded) and 850 hPa u-
734 component anomalies (contours, units of ms^{-1}) subsequent to initial large MJOs
735 (RMM amplitude > 1.0). (a) Analyzed, initial MJO in phase 1; (b) forecast, initial MJO
736 in phase 1; (c) analyzed, initial MJO in phase 4; and (d) forecast, initial MJO in phase
737 4. 133 cases were used in panels (a) and (c), 214 cases in panels (b) and (d).

738 **Figure 10:** Dec-Jan-Feb 1997-2012 daily-average precipitation climatologies for (a)
739 analyzed GPCP data, (b) +0 to 1 day forecast, (c) +7 to 8-day forecast, and (d) +15 to
740 16-day forecast.

741 **Figure 11:** Rank histograms of RMM1 and RMM2 values [green bars], the angle of
742 the vector in the (RMM1, RMM2) phase space [red bars], and the magnitude [blue
743 bars], for forecast lead times of 1 to 16 days.

744 **Figure 12:** Continuous ranked probability skill scores (CRPSS) of ensemble
745 reforecasts of the MJO relative to an unconditional climatological distribution (red
746 line) and relative to a lagged regression model using current and recent analyzed

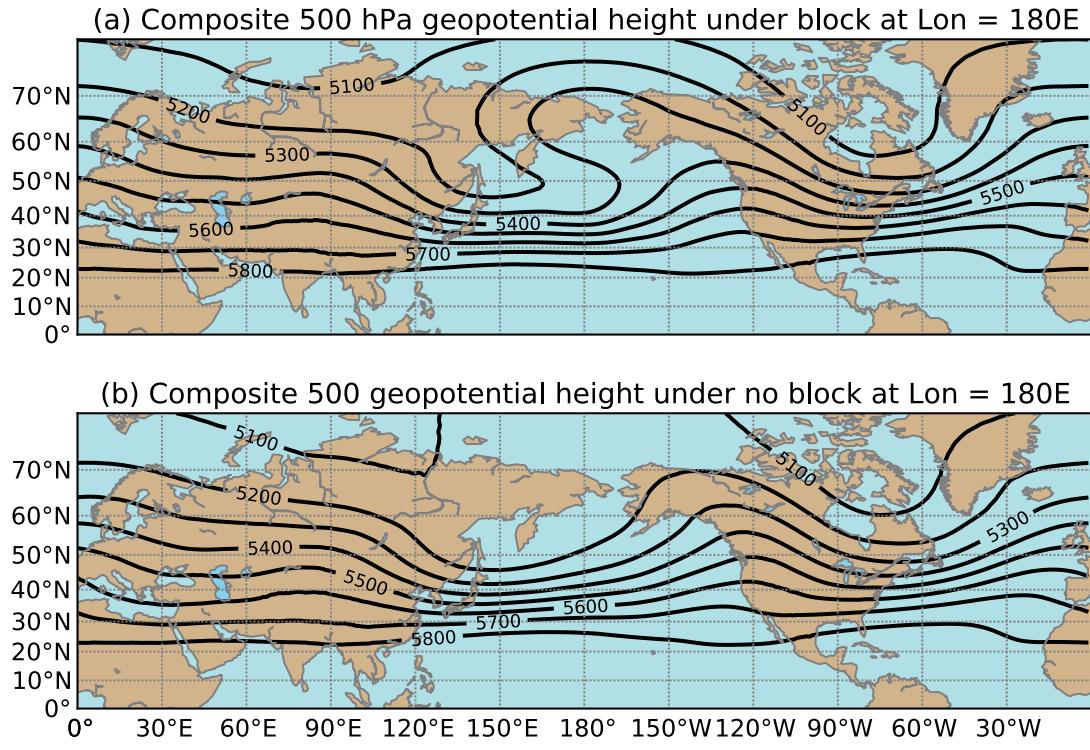
747 RMM values as predictors (blue line). The green line shows the “perfect-model”
748 skill, when one forecast member is used as a surrogate for the verification and the
749 remaining 10 members are used to generate the probabilities.

750 **Figure 13:** Correlation skill (panels (a), (c), and (e)) and RMSE (panels (b), (d), and
751 (f)) for MJO forecasts. Panels (a) and (b) show skill and RMSE for total and for
752 individual wind and OLR components of RMM. Panels (c) and (d) show overall skill
753 and RMSE for half-decadal periods. Panels (e) and (f) show skill for cases with
754 initial large and small amplitude, as defined in the text.

755 **Figure 14:** Change in blocking frequency as a function of longitude for when
756 analyzed (panel a) or +4, +8, or +16 day forecast (panels b, c, and d, respectively)
757 has RMM of phase theta + / - 22.5 degrees. Phases of MJO diagram as marked on Fig.
758 6 are noted on the right-hand side of the plots. The ranges the Euro-Atlantic and
759 Pacific Sectors are noted with heavy black lines at the bottom of each panel.

760

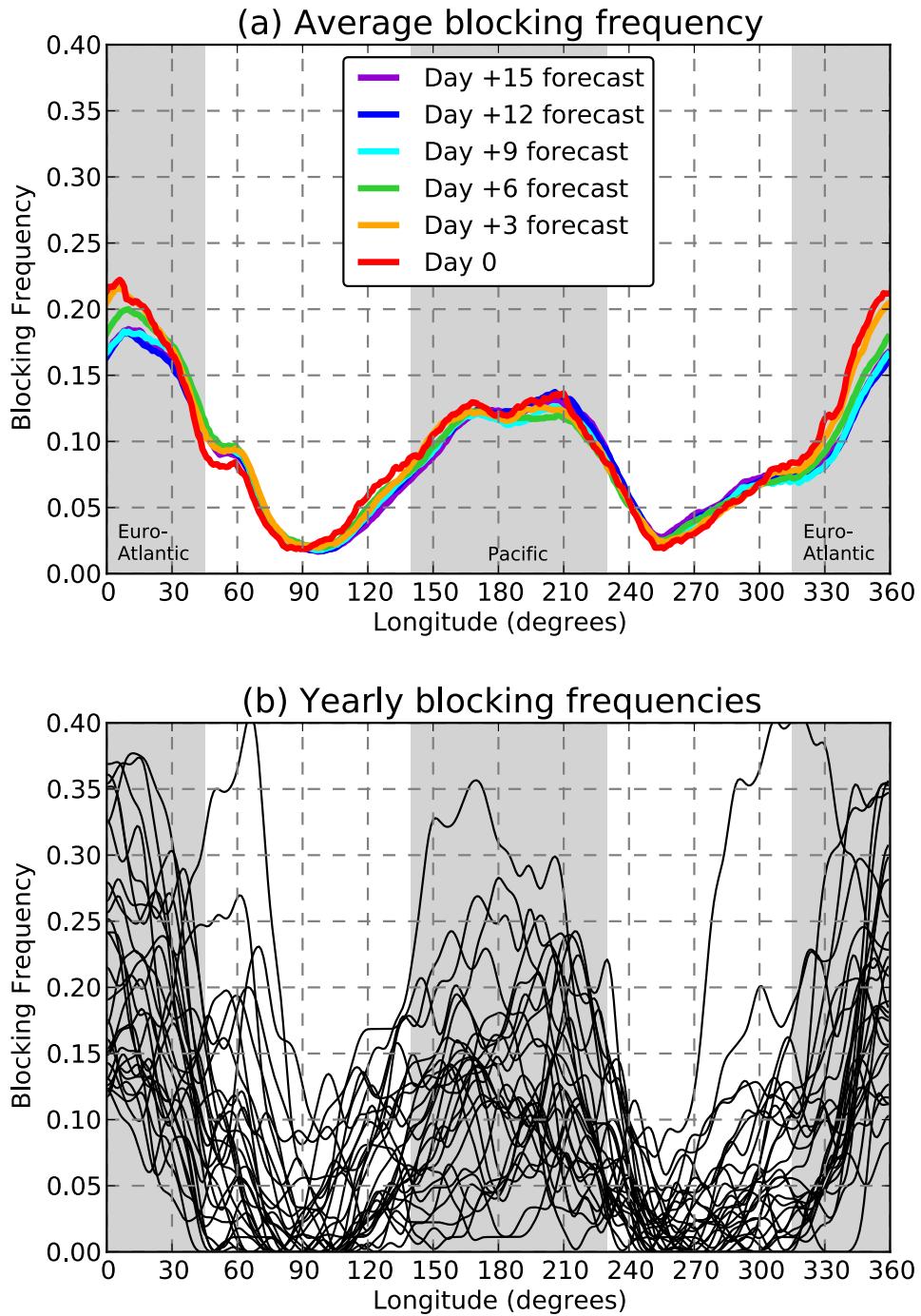
761



762

763 **Figure 1:** Composite of Northern Hemisphere 500 hPa geopotential height patterns
 764 under (a) blocked flow at 180° E longitude, and (b) unblocked flow at 180° E
 765 longitude.

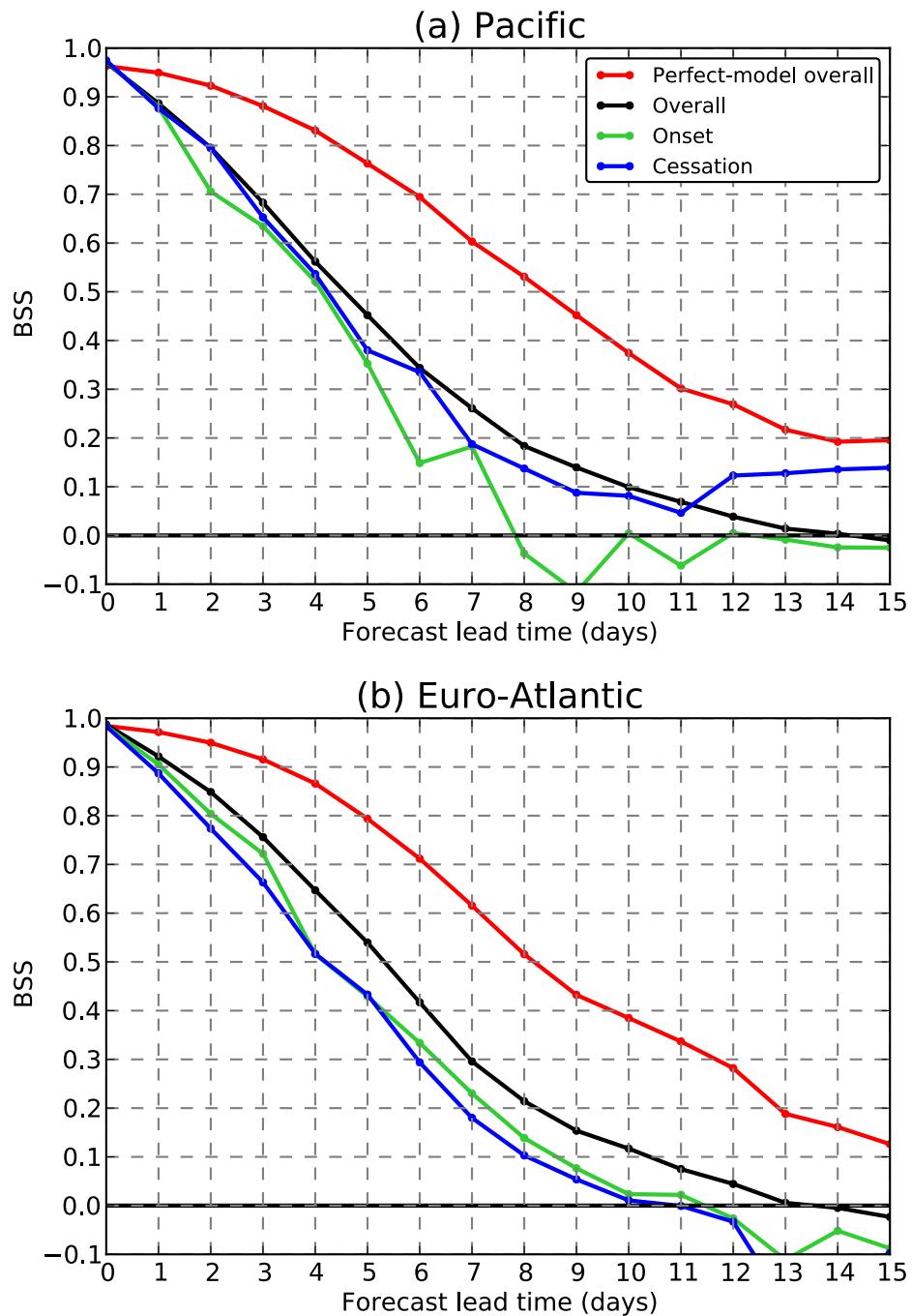
766



767

768 **Figure 2:** (a) Blocking frequency as determined from analyses and reforecasts as a
 769 function of forecast lead time. Areas shaded in gray denote the two sectors in
 770 subsequent figures, the Pacific and Euro-Atlantic sectors. (b) Analyzed, spatially
 771 smoothed yearly blocking frequencies for each year between 1985 and 2012.

772

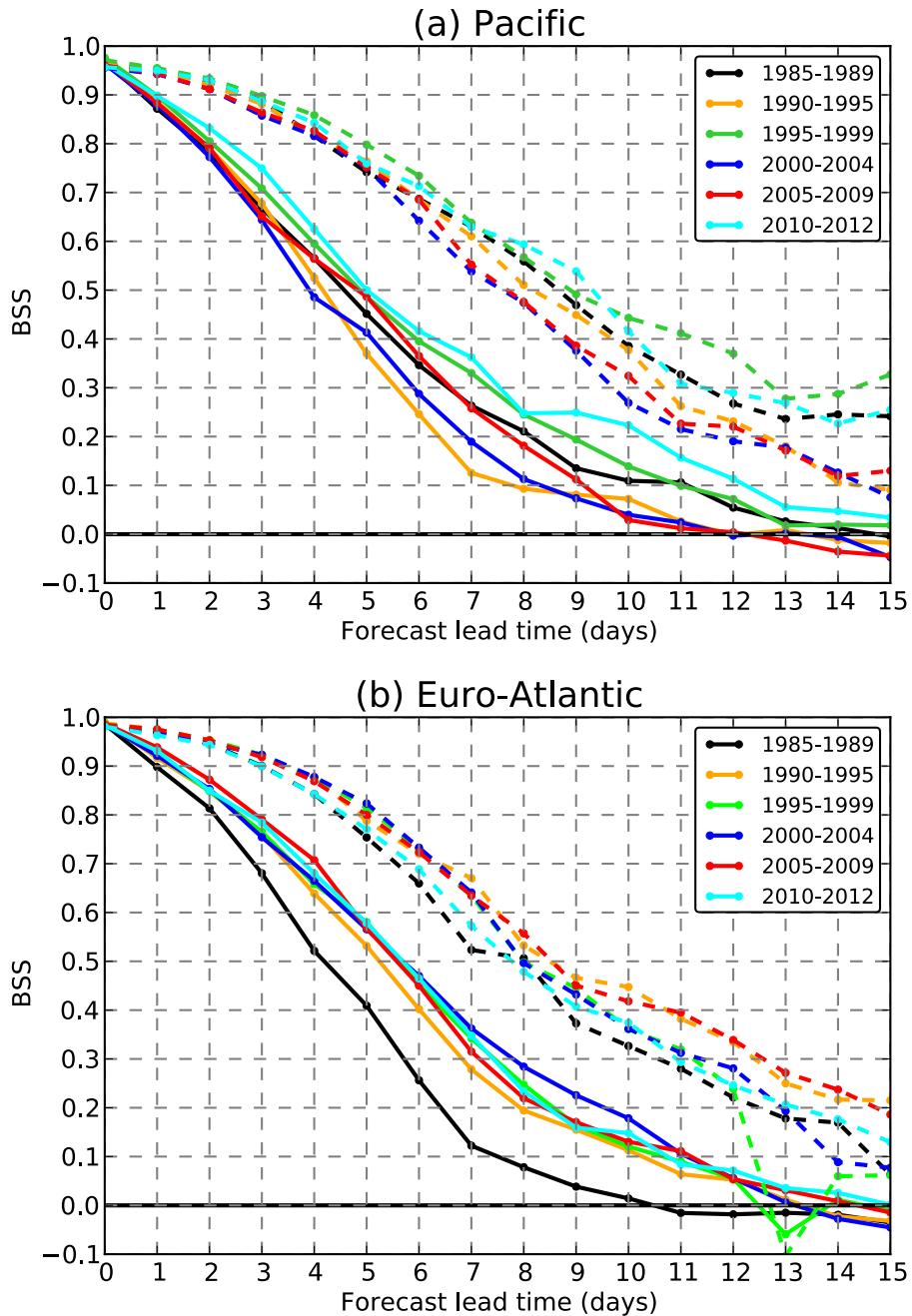


773

774 **Figure 3:** Brier Skill Scores of blocking probability forecasts for (a) Pacific, and (b)

775 Euro-Atlantic sectors.

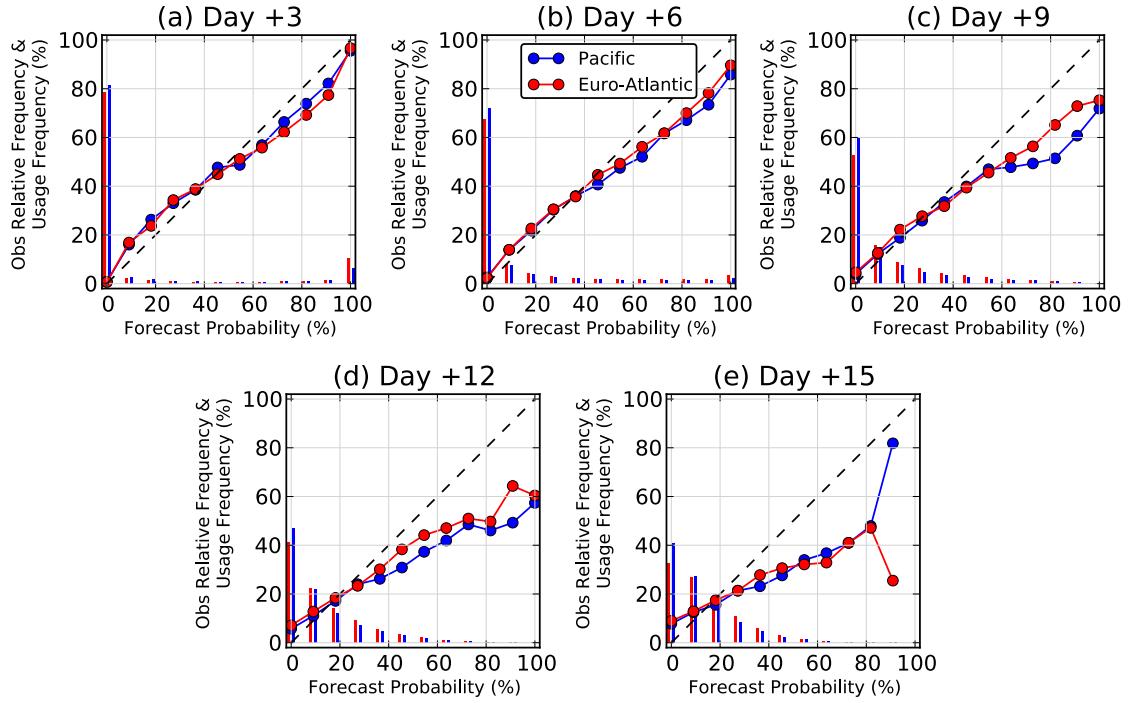
776



777

778 **Figure 4:** As in Fig. 3, but for Brier Skill Scores of blocking probability forecasts by
 779 half decade for (a) Pacific, and (b) Euro-Atlantic sectors. Solid lines present the skill
 780 scores for the actual reforecasts, dashed lines present the skill under perfect-model
 781 assumptions.

782



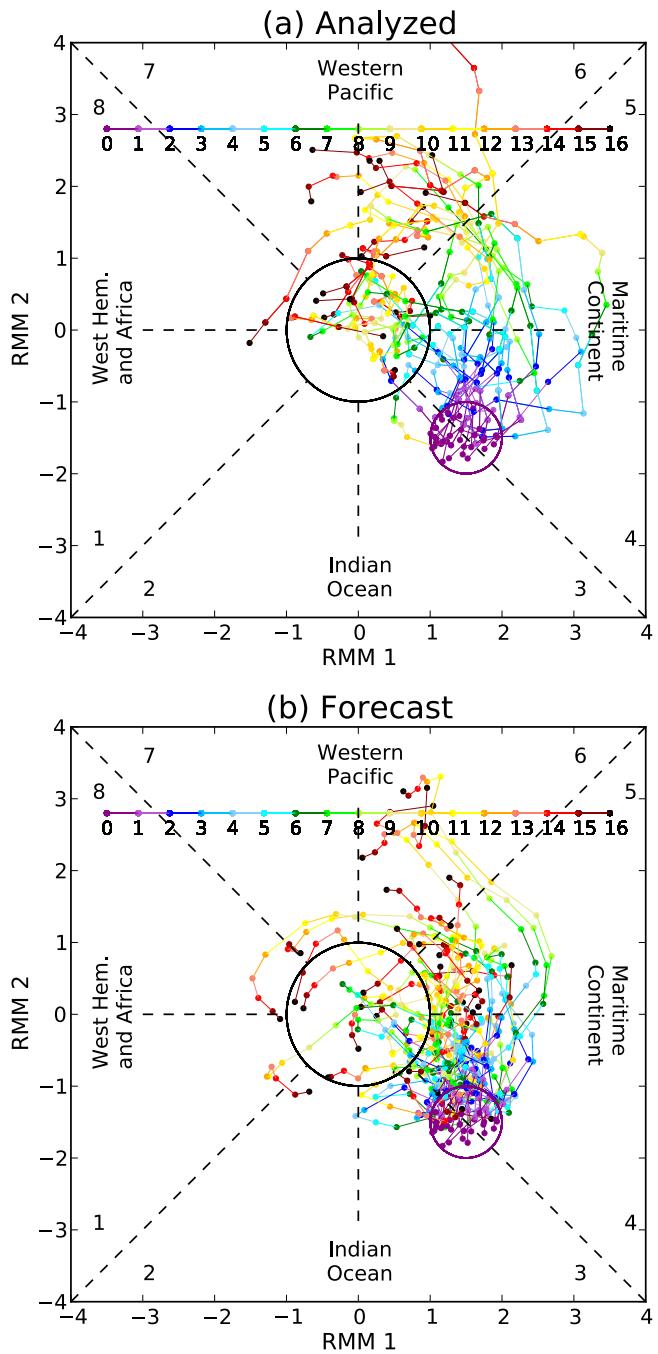
783

784

785 **Figure 5:** Reliability diagrams for blocking probability forecasts for (a) +3 day
 786 forecast, (b) +6 day forecast, (c) +9 day forecast, (d) +12 day forecast, and (e) +15
 787 day forecast. Dotted red line denotes the skill in the Euro-Atlantic sector and dotted
 788 blue line denotes the skill in the Pacific sector. Red and blue bars indicate the
 789 frequency of usage of each forecast probability category for the Euro-Atlantic and
 790 Pacific sectors, respectively.

791

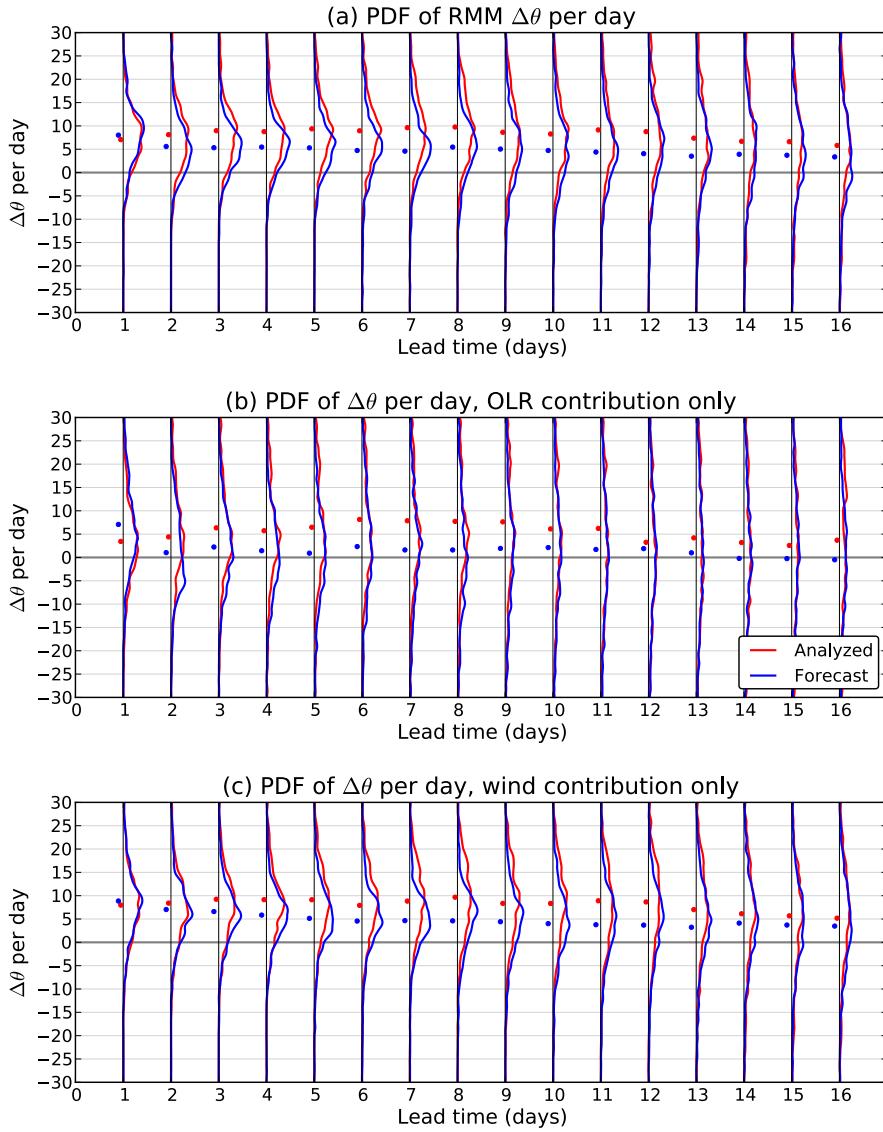
792



793

794 **Figure 6:** RMM1 and RMM2 phase plots for (a) analyzed, and (b) control forecasts
 795 whose initial states are within the purple circle. Differences in time from the initial
 796 time conveyed by the colors of lines and dots, with the legend indicating the lead
 797 time in days. Phases 1 to 8 are marked in the corners of the diagram.

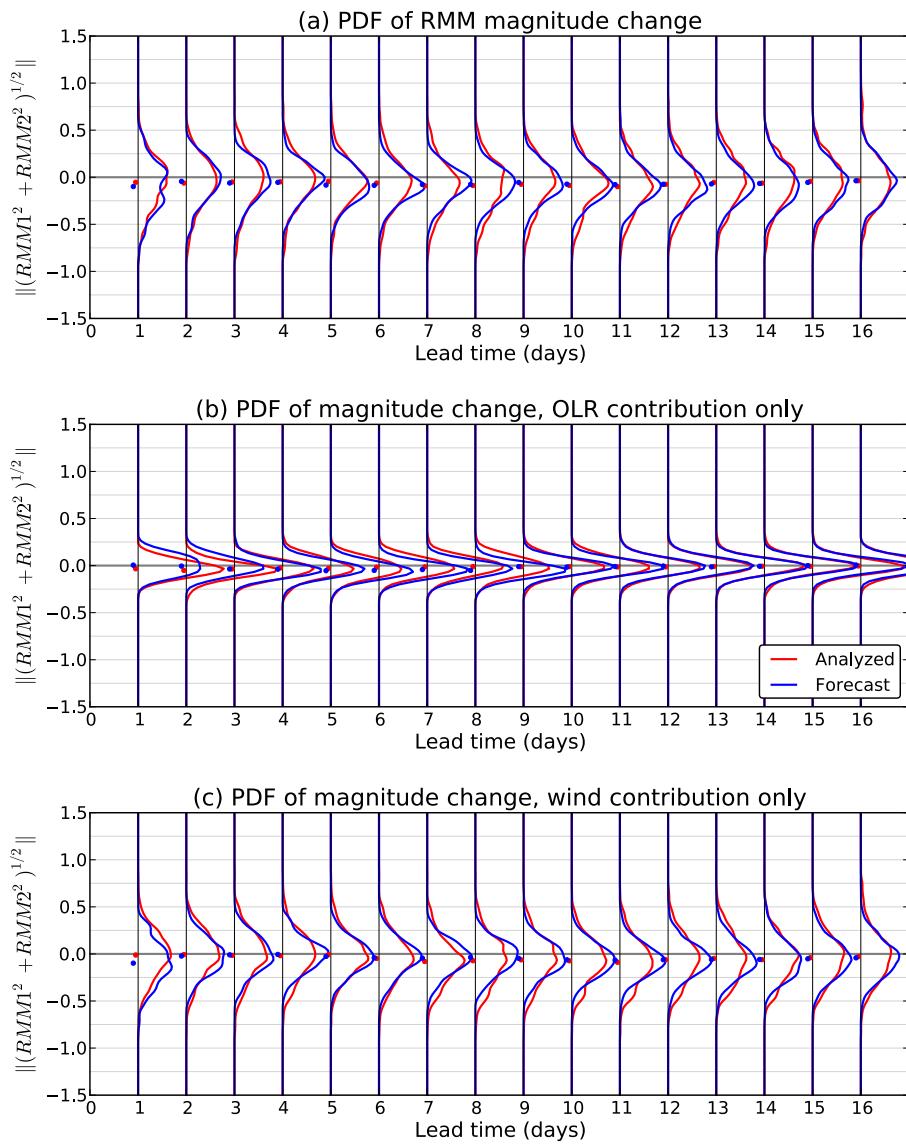
798



799

800 **Figure 7:** PDF of daily change in angle of RMM vector ($\Delta\theta$), measured in degrees, for
801 (a) overall RMM, (b) OLR contribution to RMM, and (c) wind-component
802 contribution to RMM. Only Dec-Jan-Feb 1985-2012 dates which had an initial RMM
803 magnitude of greater than 1.0, an RMM2 component < 0.0 , and an initial OLR
804 component of > 0.5 were included as samples. Dots indicate the mean of the PDF for
805 a given day, horizontally offset slightly so that dots do not overlap.

806



807

808 **Figure 8:** As in Fig. 7, but for the PDF of the change in magnitude of RMM.

809

810

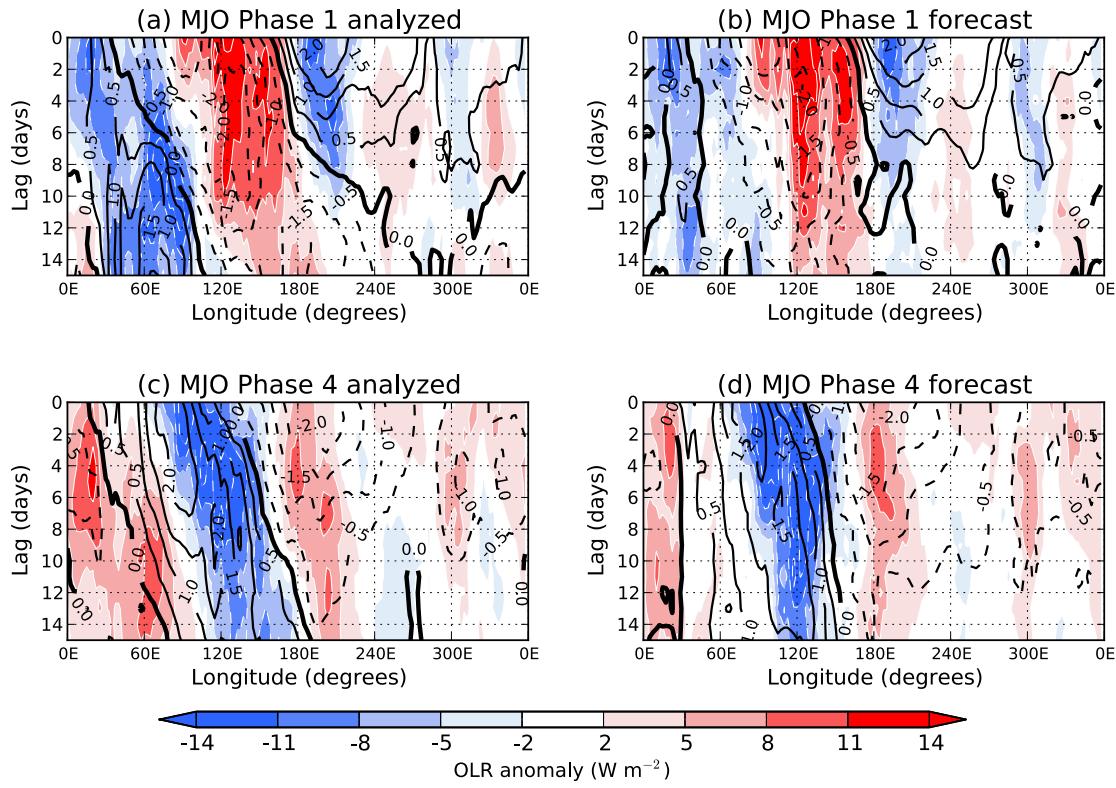
811 **Figure 9:** Composite of lagged filtered OLR anomalies (shaded) and 850 hPa u-

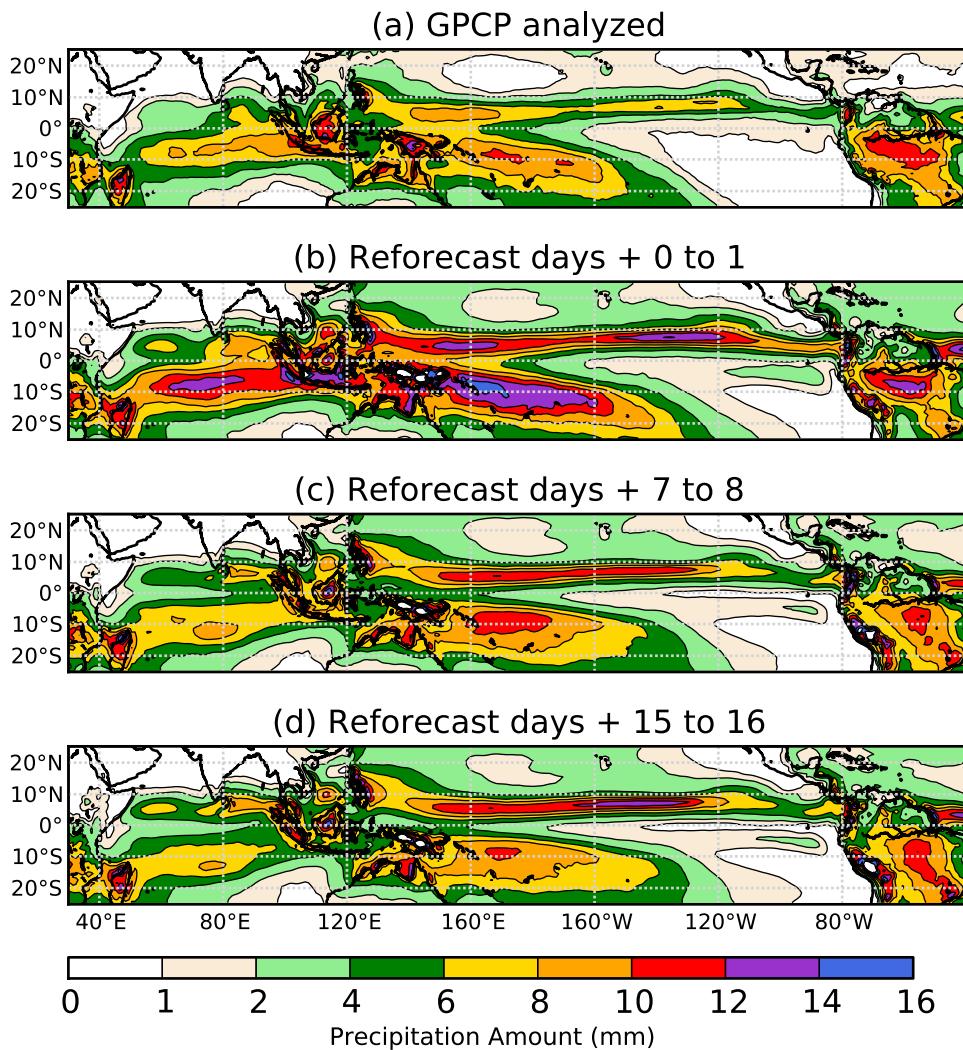
812 component anomalies (contours, units of ms^{-1}) subsequent to initial large MJOs

813 (RMM amplitude > 1.0). (a) Analyzed, initial MJO in phase 1; (b) forecast, initial MJO

814 in phase 1; (c) analyzed, initial MJO in phase 4; and (d) forecast, initial MJO in phase

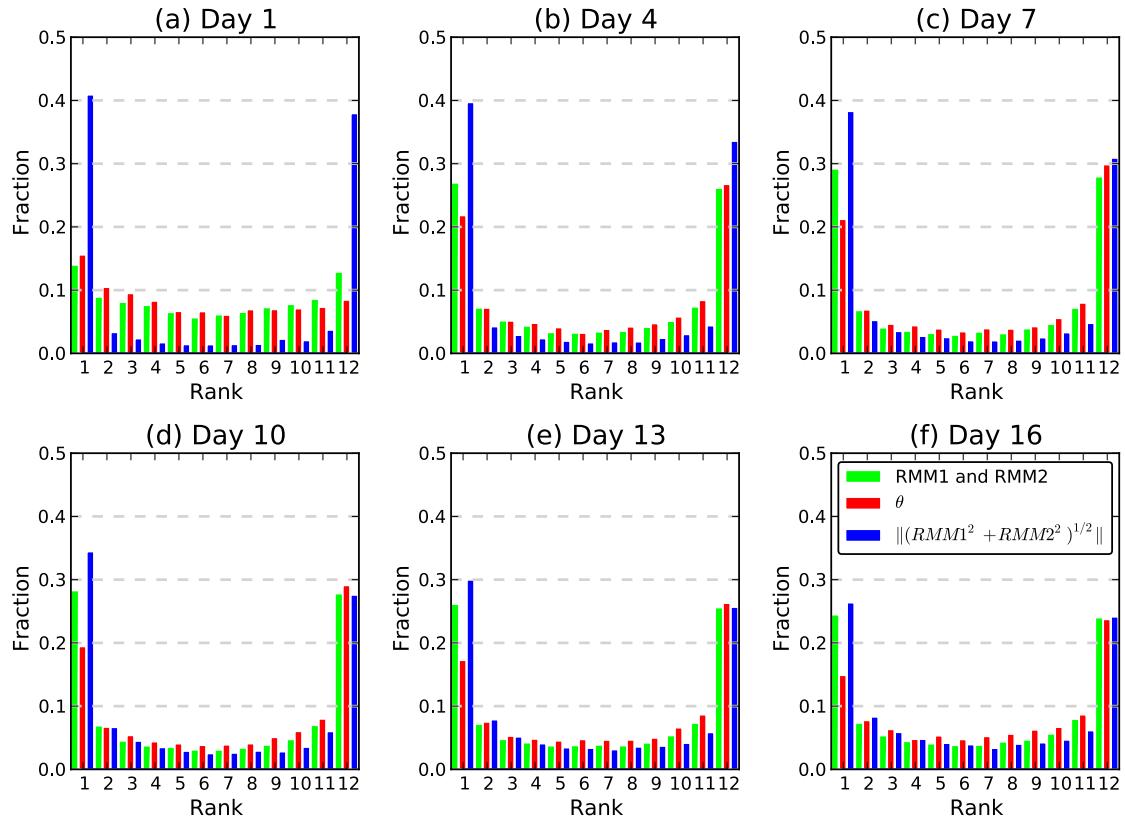
815 4. 133 cases were used in panels (a) and (c), 214 cases in panels (b) and (d).





816
817 **Figure 10:** Dec-Jan-Feb 1997-2012 daily-average precipitation climatologies for (a)
818 analyzed GPCP data, (b) +0 to 1 day forecast, (c) +7 to 8-day forecast, and (d) +15 to
819 16-day forecast.

820



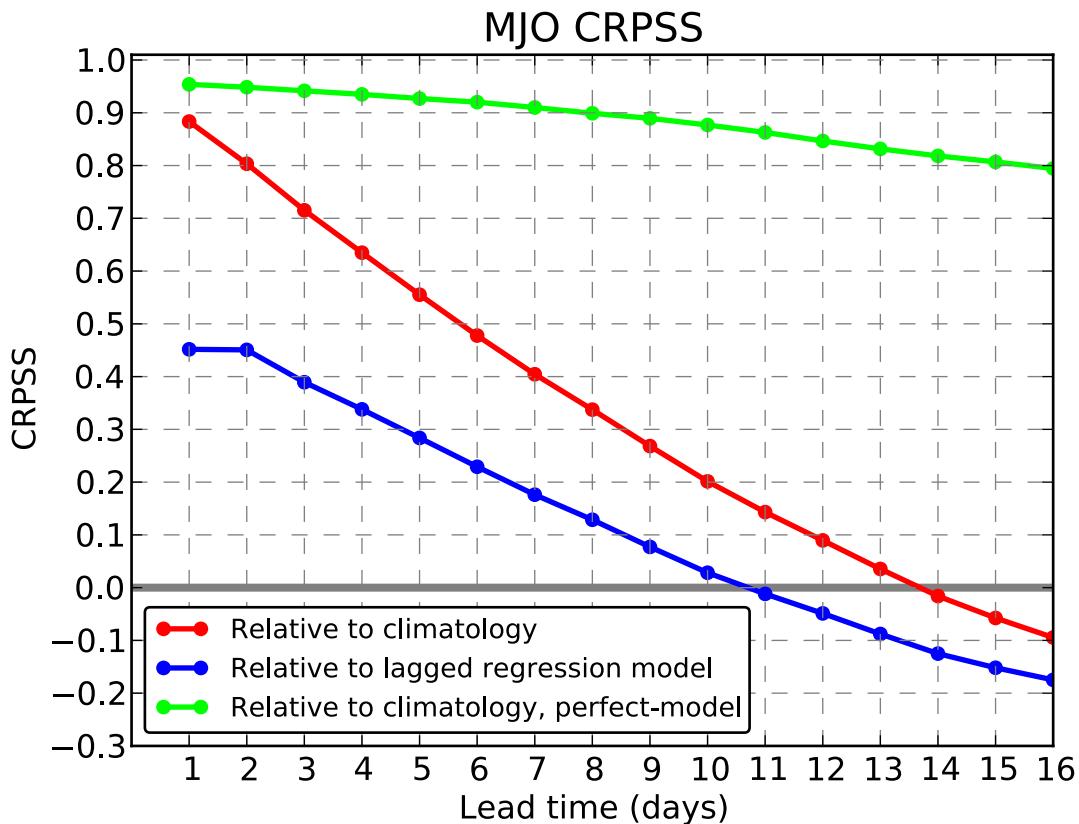
821

822 **Figure 11:** Rank histograms of RMM1 and RMM2 values [green bars], the angle of
 823 the vector in the (RMM1, RMM2) phase space [red bars], and the magnitude [blue
 824 bars], for forecast lead times of 1 to 16 days.

825

826

827



828

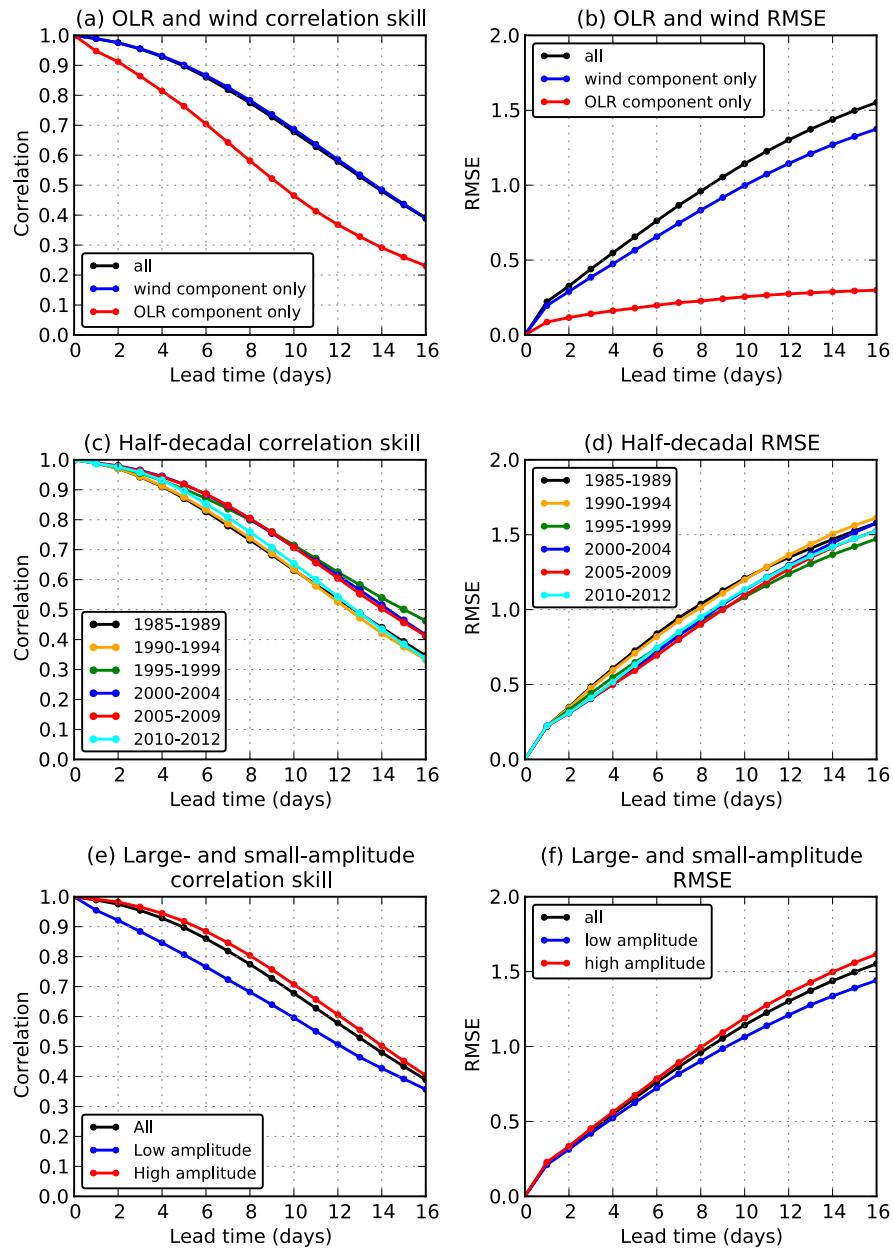
829 **Figure 12:** Continuous ranked probability skill scores (CRPSS) of ensemble
 830 reforecasts of the MJO relative to an unconditional climatological distribution (red
 831 line) and relative to a lagged regression model using current and recent analyzed
 832 RMM values as predictors (blue line). The green line shows the “perfect-model”
 833 skill, when one forecast member is used as a surrogate for the verification and the
 834 remaining 10 members are used to generate the probabilities.

835

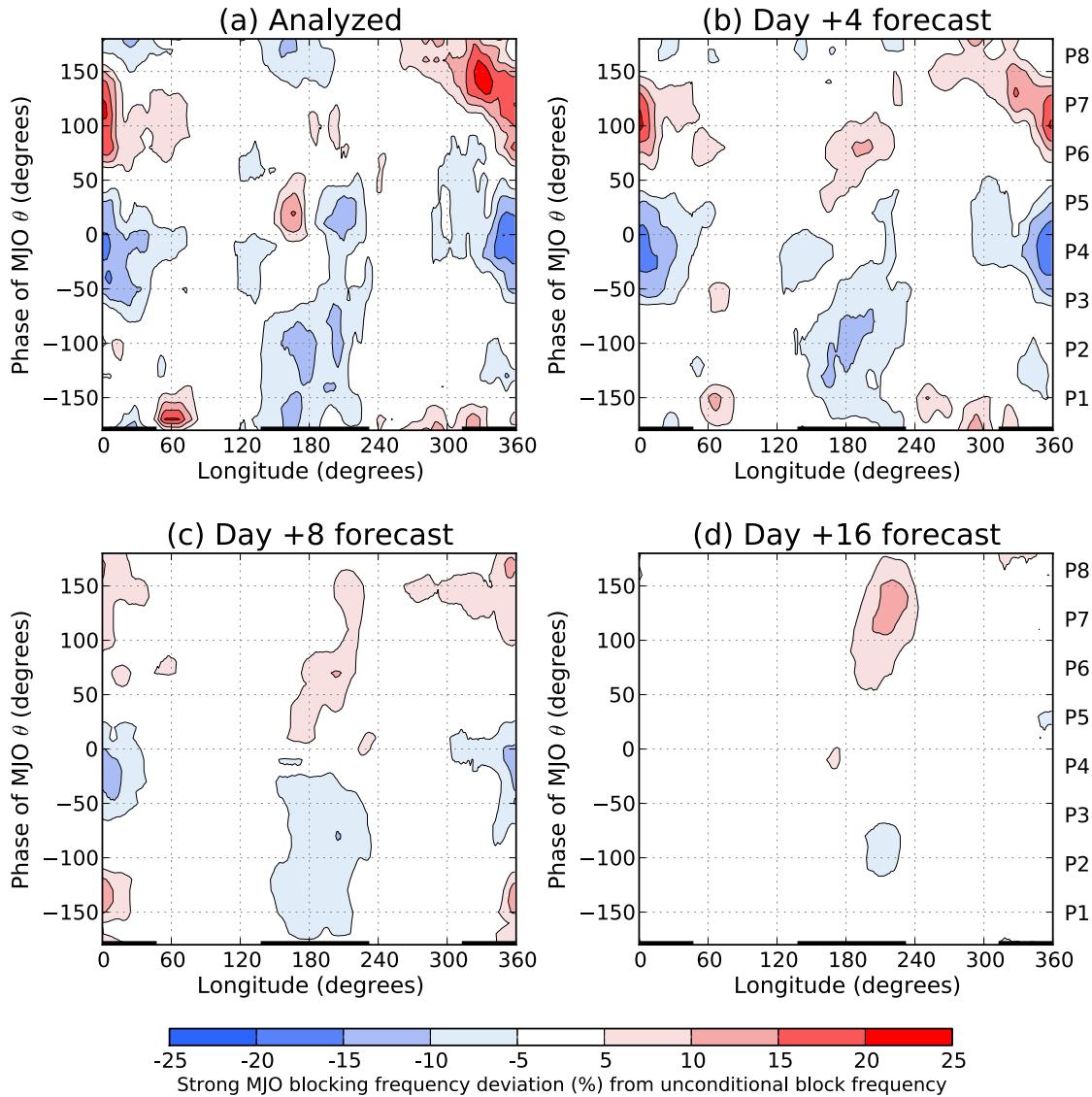
836

837

838 **Figure 13:** Correlation skill (panels (a), (c), and (e)) and RMSE (panels (b), (d), and
 839 (f)) for MJO forecasts. Panels (a) and (b) show skill and RMSE for total and for
 840 individual wind and OLR components of RMM. Panels (c) and (d) show overall skill
 841 and RMSE for half-decadal periods. Panels (e) and (f) show skill for cases with
 842 initial large and small amplitude, as defined in the text.



843



844

845

846 **Figure 14:** Change in blocking frequency as a function of longitude for when
 847 analyzed (panel a) or +4, +8, or +16 day forecast (panels b, c, and d, respectively)
 848 has RMM of phase theta + / - 22.5 degrees. Phases of MJO diagram as marked on Fig.
 849 6 are noted on the right-hand side of the plots. The ranges the Euro-Atlantic and
 850 Pacific Sectors are noted with heavy black lines at the bottom of each panel.